DOCUMENT RESUME

ED 135 827 TM 006 052

AUTHOR Nieves, Luis
TITLE The GRE and the Minority Student: A Perspective.
INSTITUTION Educational Testing Service, Princeton, N.J.
PUB DATE [Sep 76]
NOTE 29p.; Paper presented at the Annual Meeting of the American Psychological Association (84th, Washington, D.C. September 3-7, 1976)

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS *Admission Criteria; *College Admission; *College Entrance Examinations; *Graduate Study; Higher Education; *Minority Groups; Predictive Validity; Predictor Variables; *Standardized Tests; Test Bias; Testing Problems; Test Validity
IDENTIFIERS *Graduate Record Examinations

ABSTRACT

Alternate admission systems are reviewed in respect to the minority applicant in an attempt to assess the nature of the bias frequently referred to in connection with standardized testing. The connection between the use of the Graduate Record Examinations (GRE) as a standardized test and the admission system was explored, concluding that the role and importance of the GRE as an admission criterion was determined by the type of admission system adopted. The psychometric definitions of test fairness (bias) were also reviewed, noting that there are several conflicting definitions of test fairness and that these closely parallel the types of admission systems currently in use. The validity of the GRE was examined, pointing out that the validity, as with most standardized tests, varied with the validity criterion used and various other factors. The need for local departmental validity studies was stressed in addition to the need for special studies for minority groups. Of the three major sources of potential bias common to standardized testing, content bias, environmental bias, and utilization bias, the last was noted as bearing most directly on the concern for testing as a barrier to the admission of minority students into educational institutions. The question of testing as a barrier, in the last analysis, rests with the values which dominate the selected selection of admission criteria in general. These criteria, usually selected with an eye to the outcome of an educational process, are perceived as the necessary prerequisite to a successful academic experience. (Author/RC)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED135827

THE GRE AND THE MINORITY STUDENT: A PERSPECTIVE

By Luis Nieves

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY

Luis Nieves

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

TM006 052

September 6, 1976
APA Symposium:
Attracting, Admitting, and
Maintaining Minority Graduate Students

Copyright © 1976 by Educational Testing Service. All rights reserved.

THE GRE AND THE MINORITY STUDENT: A PERSPECTIVE

The Admission System - How the GRE are Used

From the time it was clear that minority populations were not proportionally represented in post secondary education many inferred that the entrance examinations were the barrier. This position was ostensibly supported by the consistent score differentials between majority and minority test takers coupled with the low number of applicants and admissions into graduate education. As in many of the standardized tests, minority students consistently scored lower as an aggregate group in admission tests. Where admission policies are intent on preserving a narrow definition of "high standards" they increasingly rely on the objective normative score as an indicater of "ability." Admission personnel in these institutions argue that the poor performance (low score) of individuals on these nationally comparative examinations indicates a lack of ability and/or a lack of preparation. Many institutions feel that admitting students who perform below the national average, or below a specified score, will be interpreted as a lowering of standards and suggestive of a lower quality education. Institutions evidently evaluate each other on the quality of their students as reflected in the careful distinctions between high and yet higher scores on nationally normed tests.

This paper will review the basic data and positions taken relevant to the concern that minority students are barred from entrance into and progress in graduate education because of the nature, the content and the use of the Graduate Record Examinations. Standardized admission testing will be used interchangeably with GRE and other specific tests to emphasize that the characteristics of standardized testing are the major issue rather than a specific examination. The review will include the author's perspective on the role of the selected aspects of the admission procedure, primarily, the admission system, the predictor variables, and the success criteria variables.

The admission problem for mainstream graduate institutions in some ways is the ideal one. There are many more applicants than there are available positions. When an institution does not have more applicants than openings, then the usefulness of the GRE in its current form is considerably diminished. If the GRE are not used as a selection predictor, they may be useful, but their secondary value as a diagnostic or evaluative tool has not been explored and consequently less understood. The GRE are designed as a selection tool, and so we will direct our comments to the GRE as a selection criterion.

When an institution has more applicants than positions, one of three broad admission procedures is generally followed with respect to minority applicants. The objective system, which claims to be color blind, simply ranks all the applicants by some criterion. Every student is "objectively" compared with students throughout the country (a norm) and then those who rank highest among the applicant group,

by single or multiple criteria, are selected. This system has yielded very skewed admission and attendance curves by racial or ethnic groups, although this procedure is still strongly defended as the only truly objective one. The defense of this procedure usually rests on the significant improvement it represents over the traditional "influence" model where primarily the rich and influential had extensive educational opportunity. It seems clear, however, that this objective model has not and may never yield parity in higher education so long as equal opportunity in education is restricted to the majority groups.

A second model of admission evaluates the two applicant pools separately by some criteria, a percentage of the available openings is allocated to each group and then the objective ranking system is applied to each group independently. The highest ranked students in each group are admitted before those with the lower ranks. This separate norms system guarantees a fixed percentage of the available openings to each group and consequently has its unequivocal advocates as well as vigorous detractors. It has not been popular, perhaps because of the complex legal and social implications of segregated admissions. The third model is the compromise position. Essentially, this model simply adds a preselected number of points or weighting to the minority applicant and then utilizes the first model; anticipating a more even selection distribution by ethnic and racial groups in the top tenths or quartiles of the score distribution.

The role of the GRE in each of these admission systems when applied to graduate education varies. In the objective ranking system the GRE could be used to limit minority admissions because the

system often assumes that the GRE scores mean the same thing for all persons regardless of educational background, and that low GRE scores predict precisely the opposite of what high scores predict. When you objectively (sometimes this means blindly) choose applicants by rank on a single or multiple weighted criterion (as GPA or GRE or combined), minority students are not proportionally represented among the first percentiles or quartiles. The utilization of this admission system assumes that all applicants have had equal education opportunity and a generally homogeneous cultural and social learning history. The system ignores the large cultural and socialization differences in the applicants by income, race and ethnicity.

The separate norms system which segregates minority from majority applicants may eliminate the impact of those variables which account for differential performance by race or ethnicity. In this system the performance of minority students is compared with other minority students competing for a predefined number of openings. Minority and majority students do not compete against each other. In spite of the admission guarantee benefits of this system, it does not enjoy popularity among either majority or minority educators. The procedures for differential admissions irritate the negative memories of segregation. Often there are, in addition, backlash consequences for the institutions which attempt this system serving to further inhibit its use.

The compensatory model is particularly popular with institutions, perhaps because it can be implemented anonymously and without prior commitment. The admission procedure for most graduate institutions is intentionally clouded in mystery because obscurity serves to maintain the "flexibility" desired and protect the independence of the

admissions process. This system is particularly popular in decentralized admissions process which need to change periodically. The GRE score, in this model, is the selection criterion to which a compensatory bonus is added or alternately the scores are retained as they are but given less weight for minority students. The degree to which the GRE would be a barrier to minority students would then depend on the weighting the GRE score receives and/or the bonus points which are added to them.

These three forms of admission systems suggest a possible answer to a popular question about the GRE; i.e., is the GRE score a barrier to minority students who desire graduate education? The answer, unfortunately, cannot be a straightforward yes or no but instead must be conditional on the admission system adopted and/or the role and interpretation given to test scores. The degree to which any criterion becomes a systematic barrier rests on the selection system adopted which in turn result from the values and mission of the institution. When the objectives of an institution include equal educational opportunity and a commitment to an admission program which supports that goal, then the GRE or any other selection criterion are rarely made a barrier to access. This position is often reflected in the controversy related to the charge that the content of standardized tests are biased against non-traditional test takers. This bias, it is charged, serves to misrepresent the abilities and achievements of minority applicants and, consequently, reduces the opportunities for admission.

The GRE: Are They Biased?

In assessing the role of the GRE within a prescribed admission system it will be helpful to look at the parallels between the admission system and various psychometric definitions of test fairness in relation to the controversy over bias.

There are two principal findings around which the issue of bias in test content revolves in reference to normative admission tests: first, that minority groups have lower average scores than the majority groups; second, that admission tests overpredict actual educational performance for minority groups students. Defenders of the normative testing process explain that inadequate school systems account for the first finding: the mean score differential between groups is perceived as an accurate reflection of a true difference in educational development. The finding that GRE scores "overpredict" the performance of minority students; i.e.,suggest that the students will do better than they actually do, can be a formidable hurdle to the group who contends that normative tests are culture-bound and, consequently, negatively biased for minorities. This overprediction, in actuality, is inevitable when two groups with such substantial mean score differences are compared (Cleary 1968). Thus the real problem is the score differential and the argument should return to that particular focus.

It is now generally accepted that the racial and economic prejudices which are part of American values create separate experiences, especially educational experiences for the poor and the ethnic/racial minorities as compared to the majority middle-class population. Further, that the educational system, the chief socializing agent, by providing different school experiences creates different patterns of socialization.

The tests developed for educational assessment have reflected the values and the practices imbedded in the dominant school system. The differential scores between groups may, as a result, reflect different levels and kinds of acculturation. The poor and the racial/ethnic minorities score lower because they experience very different school systems and, consequently, learn different skills that go unmeasured. and unappreciated by the admission process into graduate education.

The testing industry and the plaintiffs against testing apparently seem to concur that the different social learning history of the minority and majority students is reflected in the differential norm scores. They differ primarily when consideration is given to the action that should be taken to correct the imbalance in graduate education opportunity resulting from differential academic experiences and how that is reflected in standardized testing. The Association of Black Psychologists last year called for a moratorium on testing for minority students as an immediate solution. They followed the NAACP and other minority related educational organizations in taking this action. The testing industry, in response, has presented the evidence supporting test validity for both minority and majority groups and suggests that, if there is a bias, it helps minority students because test scores tend to overpredict their real achievement. Finally, the objective qualities of the test are described and it is suggested that more subjective methods would penalize minority applicants as they have in the past. The problem with this argument is that this conclusion is based on few studies and those studies involve assumptions or criteria which can be questioned.

There is little doubt that the best students (highest consistent GPA) tend to make relevant high scores on the GRE and, that relatively high scoring students on the GRE, on the average, tend to do well in traditional graduate schools. What about those very low scorers on the GRE who are not admitted? Can we infer that poor scores reflect future low grades or failure? Traditionally, precisely this inference is made. However, this is a difficult hypothesis to test because few low scorers are admitted into mainstream graduate institutions. Most of these institutions have the luxury of selecting among high scorers. There is some indication that low GRE scores need more scrutinizing for significance than high GRE scores. This may be true for all normative testing. In past admission studies, the lowest scores have tended to be excluded. Consequently, low GRE scores tend to be insufficiently scrutinized and overinterpreted.

Some academic investigators (Marston 1968) question the value of normative testing in general for graduate education because, they claim, the scores do not predict the most important criterion of success, post doctoral achievement. Others suggest (Nieves 1976) that normative testing has inappropriately relied on first-year grade point average as the validity criterion because, for minority students, these averages are lower than subsequent year GPAs. The evidence supports the contention that minority students obtain higher grades after the first year of either undergraduate or graduate programs. These low first-year grades when used as a validity criterion for minority students may be responsible for the overprediction findings commonly reported.

The GRE, as one of the major standardized educational tests, are now seeking to conduct research which will test the predictive validity of the GRE specifically for minority students. This has not been possible in the past because of the small number of minority students admitted to graduate education and the difficulties in coordinating multiple institutions in a comparative validity study.

The controversy between the test makers and minority educators convinced that the tests are negatively biased for minority students, will continue until a systematic effort is made to investigate minority performance on standardized tests. This effort will need more than the willingness of the testing industry, it will rely on the cooperation of graduate institutions.

Another source of possible bias other than content bias and utilization bias may be termed environmental bias. This refers to the atmosphere or the immediate testing environment including the characteristics of the tests, such as its speededness and the characteristics of the test takers, such as his/her familiarity with the format and style of the test items (test wiseness).

Of these three potential sources of bias, more fully discussed by Flaugher (1974), utilization is the most pertinent in a discussion of tests and minority test takers. We have previously discussed the admission system and the role tests may play from the perspective of the admission decision. It is valuable to look at the same process of test utilization from the perspective of the psychometric models of test fairness. These closely parallel the three admission systems and serve to more operationally define bias.

Flaugher (1974) reports that the traditional statistical model defines a fair test as one which would predict as accurately for one group as it did for the other. It is by this criterion that current standardized tests are found to be positively biased toward minority test takers (overprediction). It is contended by the psychometric researchers that the overprediction phenomenon would tend to give minority applicants a slight boost in their admission consideration. By the traditional definition of test fairness, bias in the test is generally accepted, but it is accepted as slight and positive for minority applicants. Defining fairness in a different way can produce altogether different results and have different implications for the type of selection system which would dominate the admission process.

Other definitions of test fairness have been proposed by other investigators who have significantly departed from the traditional definition (Flaugher 1974). These new definitions carry implicit criticisms of the current traditional model which dominates practice. For example, Thorndike (1971) proposed that the percentage rate of success be empirically determined for each group. Then that percentage is used to allocate available openings. This sytem, or a form of it, is familiar to admissions officers who use separate rankings of applicants to fill predetermined allocations of openings for each group and is identical to the admission system we have referred to as a separate norms systems. Adoption of Thorndike's definition of test fairness would justify the use of a separate norms system because fairness is defined along the separate norms of the sub-populations. In addition, Thorndike offers an objective method to allocate openings between the

two groups, i.e., percentage rates of success. Further, the use of this definition would tend to eliminate the effects of differential scores and place the burden squarely on empirical observations of success rates for each group. Although educators have often discussed the value of separate norms, research to support more definite conclusions has been lacking.

A third model defines fairness in terms of the probability of success. A test is fair according to this model if selection by performance on the test would give all applicants who would succeed, if given an opportunity, an equal chance to be selected.

Like the Thorndike definition, this one views the traditional statistical model as unfair and, in addition, replaces the equal validity prediction criterion (objective statistical method) with the probability of success criterion.

These models conceptualized by the psychometricians to define fairness functionally parallel the admissions models developed by the educators to meet the demands for equity. In each admission system there is an explicit use of tests as a predictor variable and within each use of the test there is a definition of bias or, inversely, test fairness.

The admissions policies of graduate institutions should receive careful scrutinizing when assessing the barriers to graduate admission for minority students. Alternate admission models are available, each having different implications for the role of standardized testing as well as definitions of bias for tests as a predictor variable.

Admission System	Test Use	Defition of Bias (Test Fairness)
Objective - (color blind) ranking of applicants. traditional model	Test is the criterion (or one of the criteria) for ranking	A test is unfair when it does not predict equally accurate for one group (minority) as it does for the other (majority)
Separate norms - ranking by groups	Norms are developed for each subgroup. Test is used for ranking within each group	A test is unfair when it does not select equal proportions of successful candidates from the two groups.
Compensatory model (equal probability model)	Adding a defined number of points and/or weight to the test score	A test is unfair when all applicants who would succeed do not have an equal chance of selection by the test.
Open admissions	Diagnostic and placement use	

The test use can successfully blend with any admission philosophy, any definition of "qualified," or any philosophy of merit or worth. The standardized test is an extremely flexible tool which will conform to any use its user dictates, including misuse.

The Validity of the GRE

"The value of a predictor for selecting student varies directly with the size of its correlation with the criterion" (Cronbach 1971). The value of the GRE overall and for specific groups depends on the size of the correlation with something we can all agree is success in graduate education. What we mean by success and what graduate deans, admission committees, faculty or departments may mean by success varies from region to region, school to school, department to department, etc. Some contend that success is a term heavily confounded with values and subjective evaluations.

Can the simple psychometric definition of success (validity criterion), soundly based on operational definitions, transcend the value laden subjective aspects of success? I don't think so. Someone needs to give the psychometricians and the researchers a better definition of success, a more reasonable validity criterion before the question of the validity of the GRE or any other standardized test can be studied more meaningfully.

The researchers and the psychometricians, lacking sufficient direction from the graduate education community, have frequently defined success as high grade point average in the first year of graduate education. By this empirical and operational definition, standardized tests, as represented by the SAT, GRE, LSAT and GMAT are reliably good predictors of success. Degree attainment is a less acceptable criterion to many psychometric researchers because it is too far removed from the predictor to comfortably establish that there is a relationship. There are problems with all success criteria but researchers seem to prefer the problems related with first year grade point average over the problems related with degree attainment as a validity criterion. The selection of the validity criterion as the definition of success becomes the key question in considering and evaluating the validity of standardized tests for minority students. For example, many minority educators prefer to use degree attainment as a criterion because first year GPAs for minority students are notoriously low and generally do not reflect subsequent year performance (Labin and Silberstein 1975). Graduate deans may not think highly of degree attainment as a validity criterion because what is considered more important to them is excellence, honors, etc. What

Table 1

Median Validity Coefficients# for Various Predictors and Criteria of Success in Graduate School**

Predictors	Criteria of Success: Graduate GPA	Overall Faculty Rating	Dept. Exam.	Attain Ph.D.	Time to Ph.D.
1. GRE-Verbal	.24 46	.31 27	.42 5	.18 47	.16 18
2. GRE-Quantitative	.23 43	.27 25	.27 5	.26 47	.25 18
3. GRE-Advanced	.30 25	.30 8	.48 2	.35 40	.34 18
4. GRE-Composite	.33 30	.41 8	*	.31 33	.35 18
5. Undergraduate GPA	.31 26	.37 15	*	.14 30	.23 9
6. Recommendations	*	*	*	.18 15	.23 9
7. GRE-GPA Composite (weighted)	.45 24	*	*	.40 16	.40 9

The lower number in each pair (set in smaller type) represents the number of coefficients upon which each median is based

* No data available

** Adopted from Willingham 1973

may be considered important is not whether the students graduate but how they graduate. The graduate community may be less concerned with predicting survivors of their program than predicting who will excel. Most graduate departments agree that many rejected students can, by some criterion, complete the program. Their task, they explain, is to predict who will do so exceptionally well.

When a clinical program in psychology needs to select ten students out of six hundred applicants, clearly more than ten can successfully complete the program. Any given clinical psychology department must define what a good clinician is, what a good clinical student is and then test for different predictors of those characteristics. Clinical departments need to conduct their own validity studies in order to insure that they are selecting the people who conform to their definition of "good clinical student" and "good clinician." In general, across fields the departmental study is the backbone of validity studies showing relationships between predictors and criteria.

The overall evidence on the efficacy or validity of predictors such as GPA, GRE or faculty recommendations suggests that they are only modest having median coefficients from as low as .14 to .48 across a variety of success criteria (Willingham 1973). Composite GRE and GPA have more consistent moderate coefficients in the range of .40 to .45.

The extremely narrow selection ratio characteristic of admission decisions gives even predictors with modest coefficients an ability to reduce selection error (see Table 2). This error, nevertheless, remains high. For three separate fields reported by Willingham (Chemistry, Physics and Psychology) attrition rates as a measure

of error were 49%, 64% and 74% respectively (see Table 3). Continued use of the present set of predictors with their modest coefficient is justified primarily by the incremental improvement over chance which they offer to all too narrow selection ratios in the admission decisions.

The validity of the GRE, consequently, is best judged relative to the particular institutions or program which would determine the success criteria selected. When success criteria are limited to the traditional ones, GPA, faculty ratings, degree attainment or time to degree attainment in a traditional program, then it appears that the GPA-GRE weighted composite is a good overall predictor yielding average coefficients of .40 to .45 which signify a reduction of the decision error of .15% to .17% when selecting five percent of the applicant pool. (See Table 2)

The Validity of the GRE for Minority Test Takers

Is the validity of the GRE for minority populations similar to that shown in the overall validity coefficients? We don't know. If the validity coefficient of an instrument can vary as a function of criteria used, size of sample, quality of design, major field (Willingham), it is understandable to hypothesize that validity coefficients vary by ethnic, racial or socio-economic group.

Studies are now being organized by the GRE Board to more directly test the question of GRE validity with minority groups. This question, though unanswered, has not gone unexplored. Flaugher (1973) has reviewed the validity data for minority students and concluded that by the dominant definition of equal effectiveness of predictions (ranking

Table 2

Proportion of Successful Selectees as a Function of Validity and Selectivity*

Validity	Selectivity (Percent Selected) 5%	10%	20%	30%	40%	50%
.00	.50	.50	.50	.50	.50	.50
.05	.54	.54	.53	.52	.52	.52
.10	.58	.57	.56	.55	.54	.53
.15	.63	.61	.58	.57	.56	.55
.20	.67	.64	.61	.59	.58	.56
.25	.70	.67	.64	.62	.60	.58
.30	.74	.71	.67	.64	.62	.60
.35	.78	.74	.70	.66	.64	.61
.40	.82	.78	.73	.69	.66	.63
.45	.85	.81	.75	.71	.68	.65
.50	.88	.84	.78	.74	.70	.67
.55	.91	.87	.81	.76	.72	.69
.60	.94	.90	.84	.79	.75	.70
.65	.96	.92	.87	.82	.77	.73
.70	.98	.95	.90	.85	.80	.75
.75	.99	.97	.92	.87	.82	.77
.80	1.00	.99	.95	.90	.85	.80
.85	1.00	.99	.97	.94	.88	.82
.90	1.00	1.00	.99	.97	.92	.86
.95	1.00	1.00	1.00	.99	.96	.90
1.00	1.00	1.00	1.00	1.00	1.00	1.00

*Adopted from Manning 1976

Table 3

Median Validity Coefficients# for Five Predictors of Graduate Success (variously defined*) in Nine Fields**

Predictors	Biological Science	Chemistry	Education	Engineering and Applied Science	English	Math	Physics	Psychology	Social Science
GRE-Verbal	.18 (7)	.22 (14)	.36 (15)	.29 (11)	.21 (6)	.30 (6)	.02 (6)	.19 (23)	.32 (11)
GRE-Quantitative	.27 (8)	.28 (13)	.28 (14)	.31 (10)	.06 (6)	.27 (6)	.21 (6)	.23 (22)	.32 (10)
GRE-Advanced	.26 (5)	.39 (9)	.24 (6)	.44 (7)	.43 (3)	.44 (5)	.38 (5)	.24 (17)	.46 (5)
Undergraduate GPA	.13 (2)	.27 (7)	.30 (5)	.18 (4)	.22 (4)	.19 (4)	.31 (4)	.16 (15)	.37 (6)
GRE-GPA Composite (weighted)	.35 (3)	.42 (6)	.42 (7)	.47 (4)	.56 (2)	.41 (3)	.45 (2)	.32 (4)	.40 (5)

The lower number in each pair (set in smaller type) represents the number of coefficients upon which each median is based

* In those sets of data where two criteria were included, one was selected for the purposes of this table in the following order of priority: GPA, Attain Ph.D., Dept. Test, and Faculty Rating

** Adopted from Willingham 1973

systems) standardized tests generally predict as well for minorities as they do for traditional students. The specific studies with minority populations in GRE to support this conclusion have not been conducted, so the question for the GRE remains open. Psychometric researchers report that there are very great difficulties which have delayed the progress of these studies. Essentially, what is needed is the cooperation and desire of graduate programs around the country to ask the question for their institutions; are the selection criteria in use equally valid for minority and majority students? The GRE Board and ETS have agreed to cooperate with institutional studies in addition to conducting their own investigations.

When researchers and statisticians claim equal validity of standardized tests for minority and majority populations they mean that relatively high scorers, regardless of race and ethnicity, perform better as judged by the validity criterion than lower scorers. This has been a general finding which, nevertheless, does not prove the assumed equal validity of standardized tests for majority and minority groups because of the tentative nature of the validity criteria used and other factors which contribute to error variance.

Coachability of the GRE - Can you Prepare (cram) for the GRE?

Having reviewed the role of the test makers and that of the test users, we need now to focus on the test taker.

For minority students or any population outside of the mainstream of American education, it is important to know all the avenues for admission

into the mainstream. When the GRE are perceived as a barrier to graduate admission, a natural consideration is made to methods of scaling the barrier. Short-term cramming is an obvious and perhaps first effort toward this end. Many counselors as a matter of course advise students to study intensively months before the exam. Many professors concerned with minority education volunteer their time to short-term tutoring. There are many commercial courses designed to prepare students for major standardized tests. Finally, there are innumerable commercial books which offer instruction on every aspect of a given standardized test. The serious counselor and student has to question whether these efforts have any proven worth.

The College Entrance Examination Board (1968) offered the public a statement which declared that short-term coaching and instruction had proven to have resulted in insignificant score increments in seven studies between 1950 and 1965. The studies cited varied in the populations sampled as well as the type and length of coaching. Up to 1970, these studies supported the conclusion that standardized educational achievement/ability tests were unsusceptible to short-term coaching. The issue seemed to rest until interest in coaching was rekindled in the early part of 1970 as the role of standardized testing was recognized to be critical in the admissions formula.

ETS and the various policy Boards working with ETS, such as the Graduate Record Examinations Board and College Entrance Examination Board, were concerned that if short-term coaching were found to significantly increase test scores, the validity of the tests would be

questioned. The efficacy and credibility of standardized testing rest on its ability to measure stable attributes which are presumed to develop over a long period of time. If scores can be quickly and easily raised, the ability of the tests to distinguish between applicants would be destroyed.

When the CEEB proposed to change the type of quantitative items on the SAT, a study was conducted to determine the susceptibility of the new proposed item types to short-term coaching. Evans and Pike (1973) concluded that for most types of quantitative items, gains can be made if the instruction is systematic, relevant to the material on the tests, given to students with a satisfactory quantitative level of development, and offered to motivated students.

These authors differentiated their study from previous ones which had uniformly found that short-term instruction produced no meaningful score gains by highlighting the weaknesses of the previous studies. These included the following: the subject's measured ability was already high and the students were a test-wise group; the amount of instruction provided was too short; the quality of the instructional material was poor. Other studies (Roberts & Oppenheim, 1966) used subjects whose mathematics background was too weak to benefit from short-term instruction.

The GRE Board up to 1970 did not consider conducting independent studies on coaching because of the definitive nature of the previous findings. As these were questioned, efforts were made to determine the effects of coaching for Black, Chicano, and White students in selected regions. These studies regrettably are not yet completed. Nevertheless, since the format of questions between the SAT and GRE in the quantitative than in the verbal area: the coached student

have a high level of similarity and often differ only in difficulty, it is not unreasonable to generalize the findings on their susceptibility to coaching from one to the other.

In general, the SAT Verbal has been reported to be more impervious to short-term coaching. The studies on which this conclusion is based have the weakness previously outlined. If I were advising students, I would say to them that having decided to seek coaching, the best investment of time is in the quantitative area rather than the verbal area.

The ETS policy boards overseeing the major standardized tests plan on maintaining a sustained investigation of their exams to short-term coaching susceptibility. One task of their test developers is to seek item types and formats which are impervious to coaching. When items can be found that do not appear to reflect stable learned attributes and skills, they will in all probability be abandoned.

When the probable benefits of the hurriedly organized commercial coaching effort are assessed, it may be concluded that for a majority of students such coaching will not make a difference. Score differentials between pre- and post-testing have varied from minus to plus one hundred points with, as well as without, coaching. It may be reasonable to assume that when large score differences are found, they may as well result from the normal error variance as from some short-term intervention. On the basis of the evidence and the experience, it may be concluded that when score increments can be made, they are usually low and rarely extremely high. Increments are more probable must have a reasonable base to begin with, and the instruction must be sustained and systematic for any kind of increment. When instruction takes this long and is sufficiently systematic, it approximates the more traditional course work in which stable learning takes place.

Summary and Conclusion

This paper has reviewed the alternative admission systems in respect to the minority applicant in an attempt to assess the nature of the bias frequently referred to in connection with standardized testing. The connection between the use of the GRE as a standardized test and the admission system was explored, concluding that the role and importance of the GRE as an admission criterion was determined by the type of admission system adopted.

The psychometric definitions of test fairness (bias) were also reviewed, noting that there are several conflicting definitions of test fairness and that these closely parallel the types of admission systems currently in use.

The validity of the GRE was examined, pointing out that the validity, as with most standardized tests, varied with the validity criterion used and various other factors. The need for local departmental validity studies was stressed in addition to the need for special studies for minority groups.

Of the three major sources of potential bias common to standardized testing, content bias, environmental bias and utilization bias, the last was noted as bearing most directly on the concern for testing as a barrier to the admission of minority students into educational institutions.

The question of testing as a barrier, in the last analysis, rests with the values which dominate the selection of admission criteria in general. These criteria, usually selected with an eye to the outcome of an educational process, are perceived as the necessary prerequisite to a successful academic experience. When City University abandoned these assumed prerequisits many educators forecasted disaster

for the students and the institutions. The reality, assessed four years after the initiation of open admission, proved that the new students did not significantly differ from the traditional student in various outcome measures (Labin and Silberstein 1975).

These open admission students did, nevertheless, demonstrate a need for a different educational process; i.e., more experiential and less didactic. If the experience of the City University can be generalized then one may conclude that the educational process for the poor and ethnic/racial minorities may need to be different, at least for the time being. Those institutions which accept the challenge and modify the traditional process so that it includes additional routes to achievement are least likely to raise admission barriers which effect the poor and minorities in a disproportionate way.

References

Cleary, T. A. Test bias: Prediction of grades of negro and white students in integrated colleges. Journal of Educational Measurement, 1968, 5, 115-124.

College Entrance Examination Board. Effects of Coaching on Scholastic Aptitude Test Scores: Statement by College Board Trustees. New York: College Board, 1975, 1968.

Cronbach, Lee J. Test Validation in Educational Measurement, R. L. Thorndike (2nd. Ed.) Washington, D.C.: American Council on Education, 1971.

Evans, F. R. & Pike, L. W. The effects of instruction for three mathematics item formats. Journal of Educational Measurement, 1973, 10, 257-72.

Flaugher, R. L. Testing practices, minority groups and higher education: A review and discussion of the research. Research Bulletin 70-41. Princeton, N.J.: Educational Testing Service, 1970.

Flaugher, R. L. Bias in testing: A review and discussion. ERIC. TM Report 36. Princeton, N.J.: Educational Testing Service, December 1974.

Flaugher, R. L. The new definition of test fairness in selection: Developments and implications. GRE Board Research Report, No. No. 72-4R. Princeton, N.J.: Educational Testing Service, May 1974.

Labin, D. E. & Silberstein, R. A. Effects of new admissions criteria upon the ethnic composition at the City University of New York. Unpublished paper, City University of New York, December 1975.

Lannholm, G. V. Review of studies employing GRE scores in predicting success in graduate school, 1952-1967. GRE Special Report #68-3. Princeton, N.J.: Educational Testing Service, 1968.

Lannholm, G. V., Marco, G. L. & Schrader, W. B. Cooperative studies of predicting graduate school success. GRE Special Report #68-3. Princeton, N.J.: Educational Testing Service, 1968

Lannholm, G. V. Summaries of GRE validity studies, 1966-1970. GRE Special Report #72-1. Princeton, N.J.: Educational Testing Service, 1972.

Manning, W. H. Some current controversies in evaluation. Unpublished report, Educational Testing Service, 1976.

Marston, A. R. It is time to reconsider the Graduate Record Examinations. American Psychologist, 1971, 26, 7.

Nieves, L. Academic performance of minorities in higher education. A paper presented at the NSSFNS Conference on Minorities' Right to Post-Secondary Education, Houston, Texas, 1976.

Roberts, S. O. & Oppenheim, D. B. The effect of special instruction upon test performance of high school students in Tennessee. Research Bulletin 66-36. Princeton, N.J.: Educational Testing Service, 1966.

Willingham, W. Predicting success in graduate education. A paper presented at the Graduate Record Examinations Board's Research Seminar at the 12th Annual Meeting of the Council of Graduate Schools. Princeton, N.J.: Educational Testing Service, 1973.